April 24, 2017

XSD-SDM Strategy/Goals Document

Strategy

The XSD Scientific Software Engineering & Data Management Group (SDM) provides leadership and scientific software engineering expertise in the areas of data management, data analysis, high-performance computing, visualization, and workflow and orchestration applications in support of world-class photon sciences at the APS. This mission is realized through the creation of a core software application portfolio in prioritized areas, including coherence, imaging, and high-energy techniques, as well with software tools for data access and management, and data streaming for real-time feedback. Effort is aligned with facility priorities and strategies, which at this time include scientific software and data management tools critical to the techniques enabled by the APS-U. Well-formed collaborative teams of software engineers, computer scientists, algorithm developers, beamline staff, users, peer groups, and other facilities and institutions develop the SDM group's software portfolio. The group works closely with the XSD-CXS group to implement new algorithms and mathematical methods, and with the XSD-BC group to integrate with beamline data acquisition systems.

Five-year Goals

The SDM group's goals over the next five years are directed at creating and deploying software tools enabling the full benefit of the portfolio of anticipated future beamlines, including the APS-U beamlines.

1. Creation and deployment of a robust set of high-performance computing (HPC) enabled software tools that address cross-cutting critical technique domain areas needed by future beamlines. This includes software in the areas of coherence, imaging, high-energy, and multi-modal techniques.
2. Deployment of a standard set of data management and distribution tools at XSD beamlines.
3. Integration of general purpose data streaming, feedback, and verification tools with beamline control software and HPC data analysis software.

Goals and Action Plan for FY2017 & FY2018

Project	Summary	FY17 SDM FTE	FY18 SDM FTE
MONA	Project aimed at integrating BlueSky with APS HPC data analysis packages and other tools to enable automated experiment feedback.	0.5	0.5
Bragg Coherent Diffraction Imaging (CDI) Software	Implement and parallelize genetic algorithms and phase retrieval methods for the Bragg CDI technique.	0.5	0.5
Correlation Toolkit	Develop a real-time HPC-enabled set of tools for time-based correlation data analysis.	0.75	0.75
General-Purpose Reciprocal-Space Mapping (RSM) Tools	Continue development and deployment of high-performance RSM tools.	1.0	1.0
X-ray Fluorescence Mapping (XFM) Software	Develop HPC-enabled fitting library and tools for fast elemental mapping.	0.75	0.75

Support for Ptychography Software	Provide ongoing support for ptychography reconstruction software and tools, and integration with complementary techniques.	0.5	0.5
Data Quality & Feedback Tools	Toolkit and framework to verify quality of collected data and provide feedback during and after acquisition.	0.5	0.5
Workflow & Data Management Tools	Continue application of analysis workflow, web portals, and data management and distribution tools at APS beamlines.	0.5	0.5
Beamline Data Streaming Prototype	Prototype software and infrastructure for real-time acquisition, streaming, and analysis, needed for APS-U and future beamlines.	0.5	0.5
Real-time Feedback & Data Acquisition System for APS-U Accelerator	Software framework and tools for the collection of data used for controls, statistics and diagnostics of technical systems for the MBA accelerator.	0.75*	0.75*
Component Database (CDB) for APS-U	An electronic system for tracking and documenting accelerator and beamline components.	0.75*	0.75*
Visualization Tools	Application and/or development of advanced visualization tools for APS beamline data analysis and experiment feedback.	0.0	0.0
Coherent Surface Scattering Imaging (CSSI) Software	Implementation of high-performance CSSI and GISAXS software applications.	0.0	0.0
Multi-modal XRF Tomography	Develop robust near real-time software for XRF tomography once algorithm development is complete.	0.0	0.0
Multi-modal Diffraction Tomography	Develop robust near real-time software for diffraction tomography once algorithm development is complete.	0.0	0.0
Multi-modal XRF Ptychography	Develop robust near real-time software for XRF ptychography once algorithm development is complete.	0.0	0.0
Laue Diffraction	Develop high-performance computing tool kit for the new Laue depth reconstruction algorithm.	0.0	0.0

* Funded by APS-U

Requirements for APS-U era

Project	APS-U Beamlines
Bragg Coherent Diffraction Imaging (CDI) Software	Atomic, InSitu, Ptycho, CHEX
Correlation Toolkit	CSSI, XPCS, CHEX
General-Purpose Reciprocal-Space Mapping (RSM) Tools	3DMicroNano, XPCS, HEXM, CHEX
X-ray Fluorescence Mapping (XFM) Software	3DMicroNano, InSitu, Ptycho
Ptychography Software	InSitu, Ptycho, CHEX
Data Quality & Feedback Tools	All
Workflow & Data Management Tools	All
Real-time Streaming Infrastructure	All
Visualization Tools	Many
Coherent Surface Scattering Imaging (CSSI) Software	CSSI

Multi-modal XRF Tomography	Ptycho
Multi-modal Diffraction Tomography	HEXM
Multi-modal XRF Ptychography	InSitu, Ptycho
New Laue Diffraction Modes	3DMicroNano

SWOT Analysis for Scientific Software

Strengths	Weaknesses
• World-leading software efforts in a number of scientific areas. • World-class beamline staff and user groups contribute new algorithms and software that expand the scientific productivity of the APS. • Highly-productive internal group of professional scientific software engineers. • Close collaborations with APS users and staff, and the XSD-CXS group to provide algorithms and with the XSD-BC group to provide integration with beamline workflows.	• Current funding situation does not allow for the APS to meet its entire mission-critical data analysis software needs. • Most current generation data analysis tools are not suited to stream data in HPC environments needed to keep up with anticipated data rates. • Many scientist-developed packages lack professional software engineering needed to make them more productive • Lower facility productivity due to lack of data analysis tools.
Opportunities	**Threats**
• Collaborations with ANL expertise will help bring state-of-the-art HPC applications to the APS. • Collaborations with DOE facilities and resources could amplify development efforts, and provide needed software in a cost effective manner for the entire DOE complex. • The APS Upgrade-enabled techniques may be fully realized, answering new scientific questions; the APS maintains its position as the most productive light source.	• Without further investment and collaboration in this area, the APS will not fully realize the scientific potential of the APS Upgrade. • User groups may seek to perform cutting-edge experiments at other light sources where better software support is available. • Other domestic and international light sources have considerably larger and more active software and algorithm development programs that can leapfrog APS leadership.

SWOT Analysis for Data Management & Distribution

Strengths	Weaknesses
• World-leading expertise at ANL in data sciences, data management and transfer (e.g. Globus Services team). • APS is the DOE's largest data collecting user facility, producing a wealth of scientifically valuable data. • Collaborative efforts continue to form between the APS and expertise elsewhere at ANL.	• Preponderance of existing unique solutions at beamlines involving manual, inefficient management steps; no common user experience. • Current manual methods cannot keep pace with increasing data rates. • Lowered productivity due to time taken away from staff and users to address tasks that may be automated.
Opportunities	**Threats**
• Leverage expertise from CLS, UoC, and the Globus Services team. • Reduce cost by leveraging outside software resources and expertise. • Consistent data management user experience. • Increase scientific productivity through automation of data management tasks.	• The full potential of the APS Upgrade cannot be realized without managed data workflows. • Lowered scientific productivity due to an inability to keep up with increases in data. • International light sources that have invested heavily in data management software may overtake the APS in terms of scientific productivity.